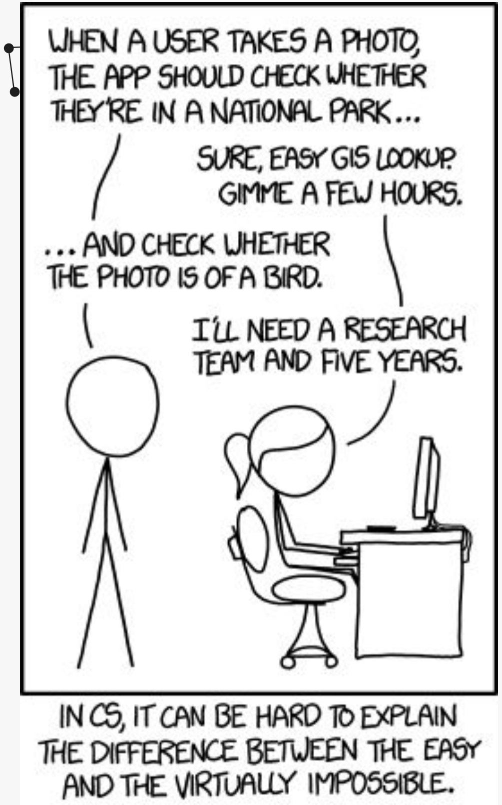
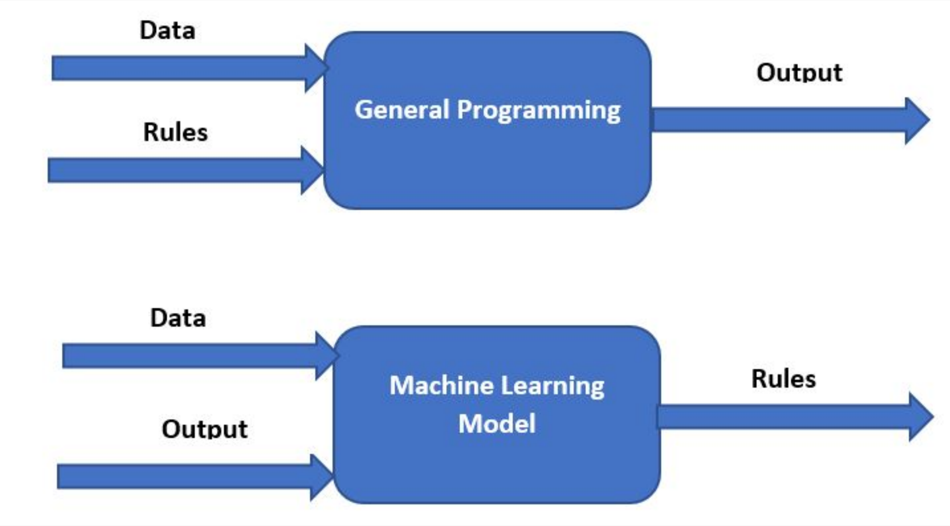
Introduction to ML

Invento Research Inc.

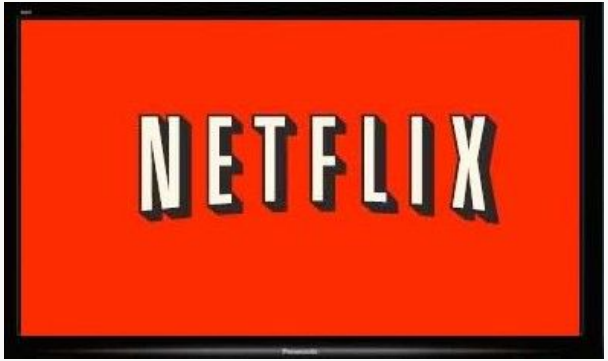
What is Machine Learning

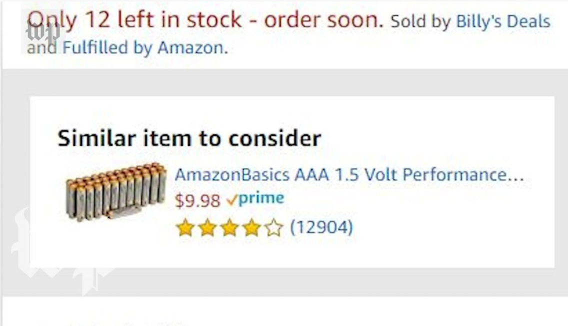


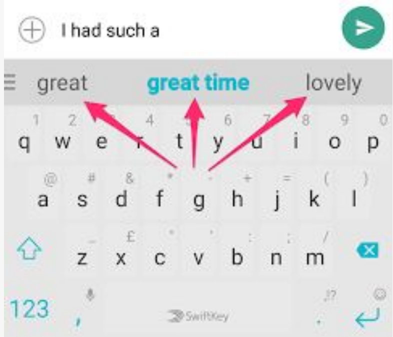


Software 2.0

ML examples


NETFLIX

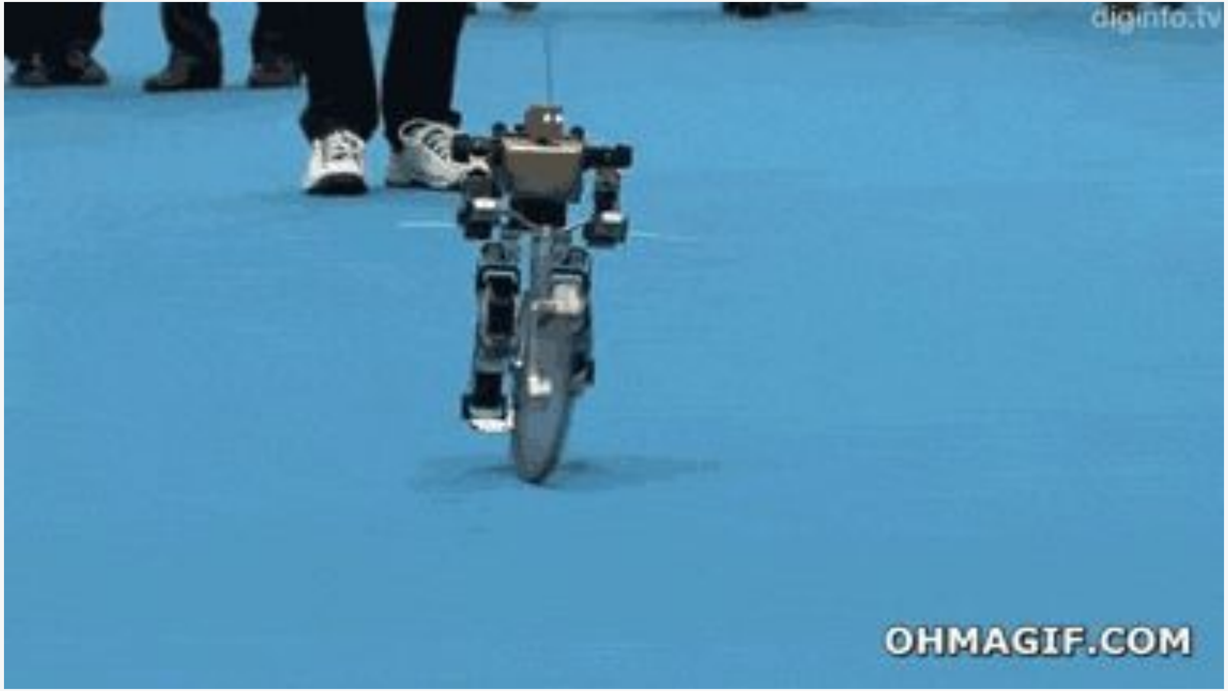

Only 12 left in stock - order soon. Sold by Billy's Deals
and Fulfilled by Amazon.
Similar item to consider
AmazonBasics AAA 1.5 Volt Performance...
$9.98 prime
(12904)


I had such a
great
great time
lovely


Followers
Instagram

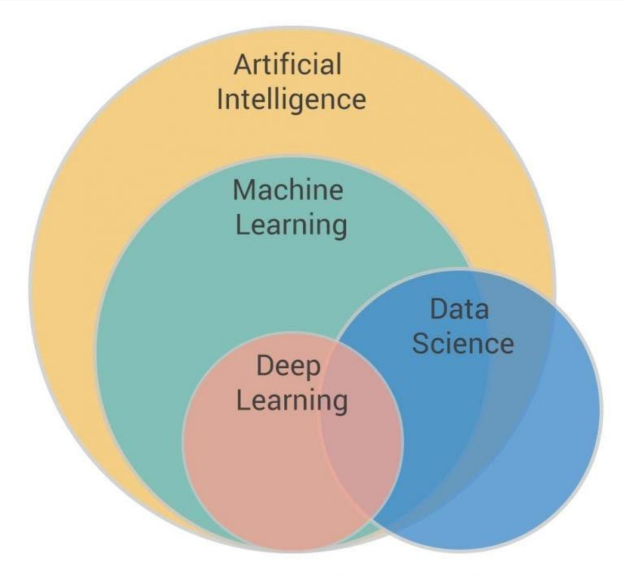

$#*%!
#ToxicTwitter

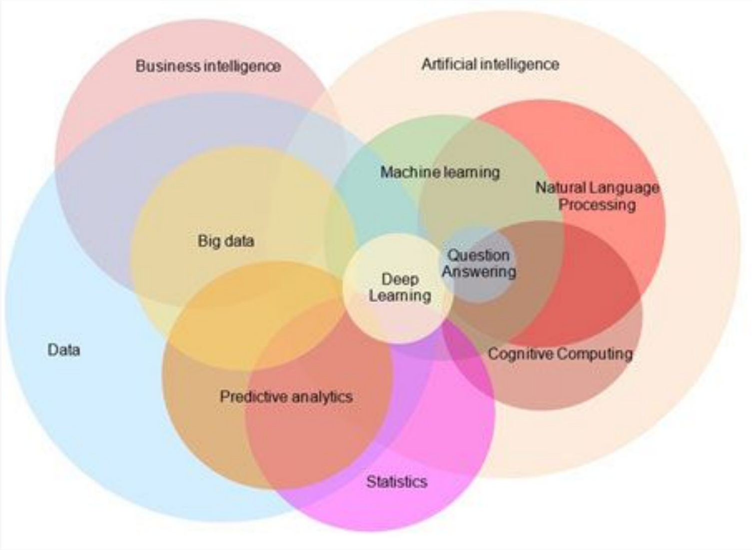
Why Machine Learning?



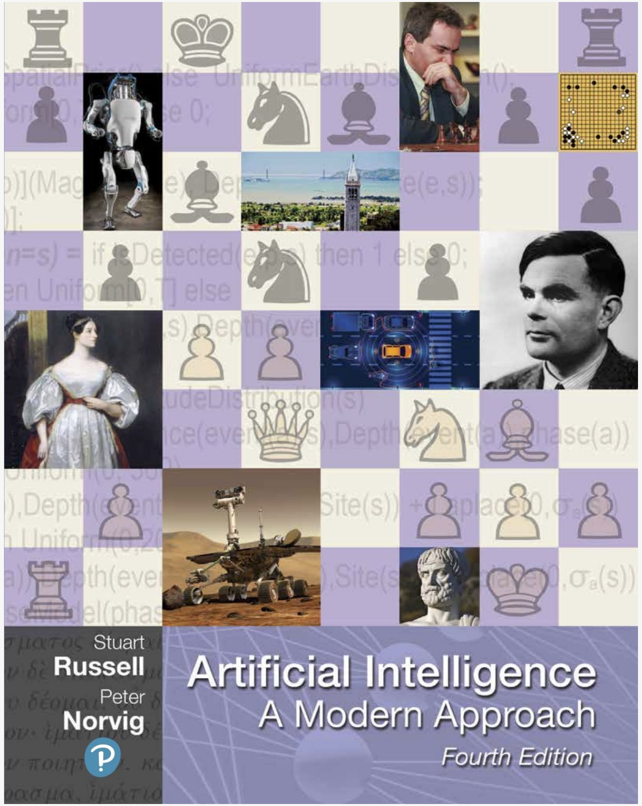
We don't know how to program some things!

Venn diagram of ML and deep learning



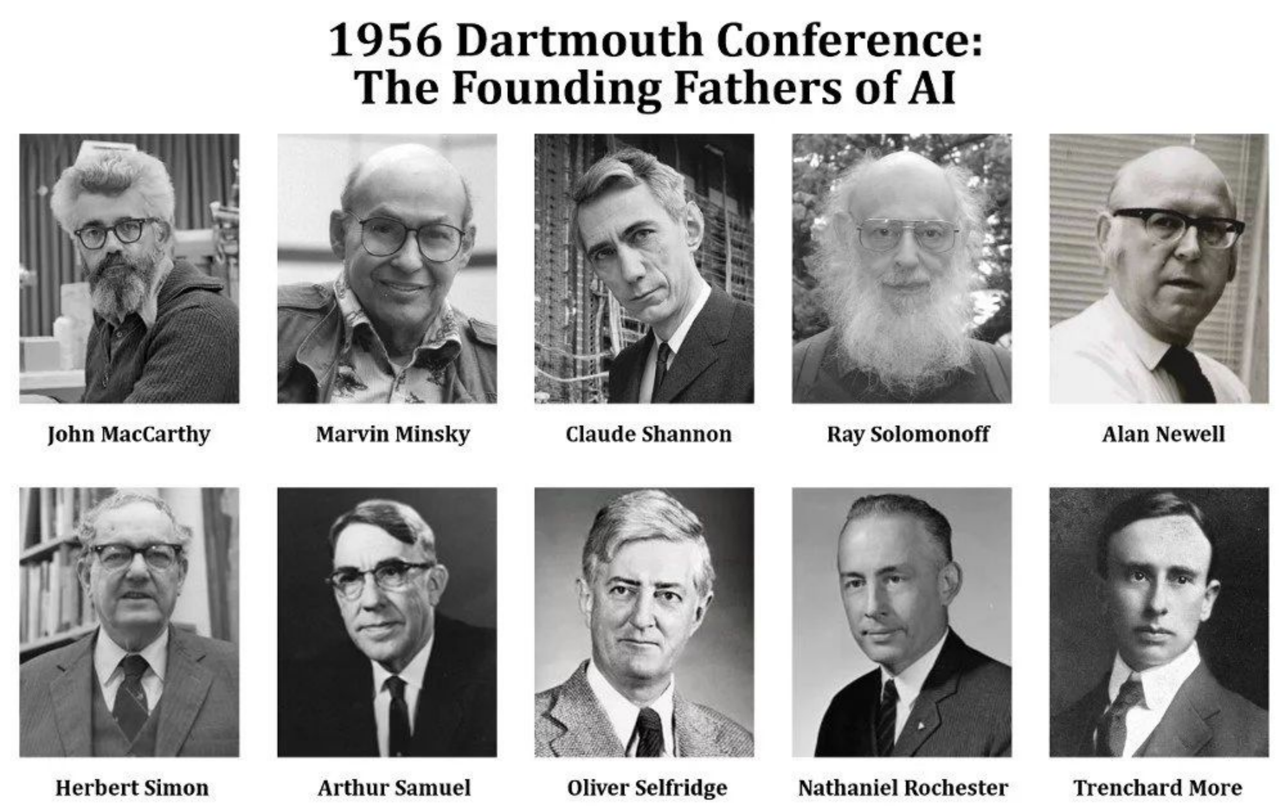


Gurus of GOFAI



Stuart Russell and Peter Norvig

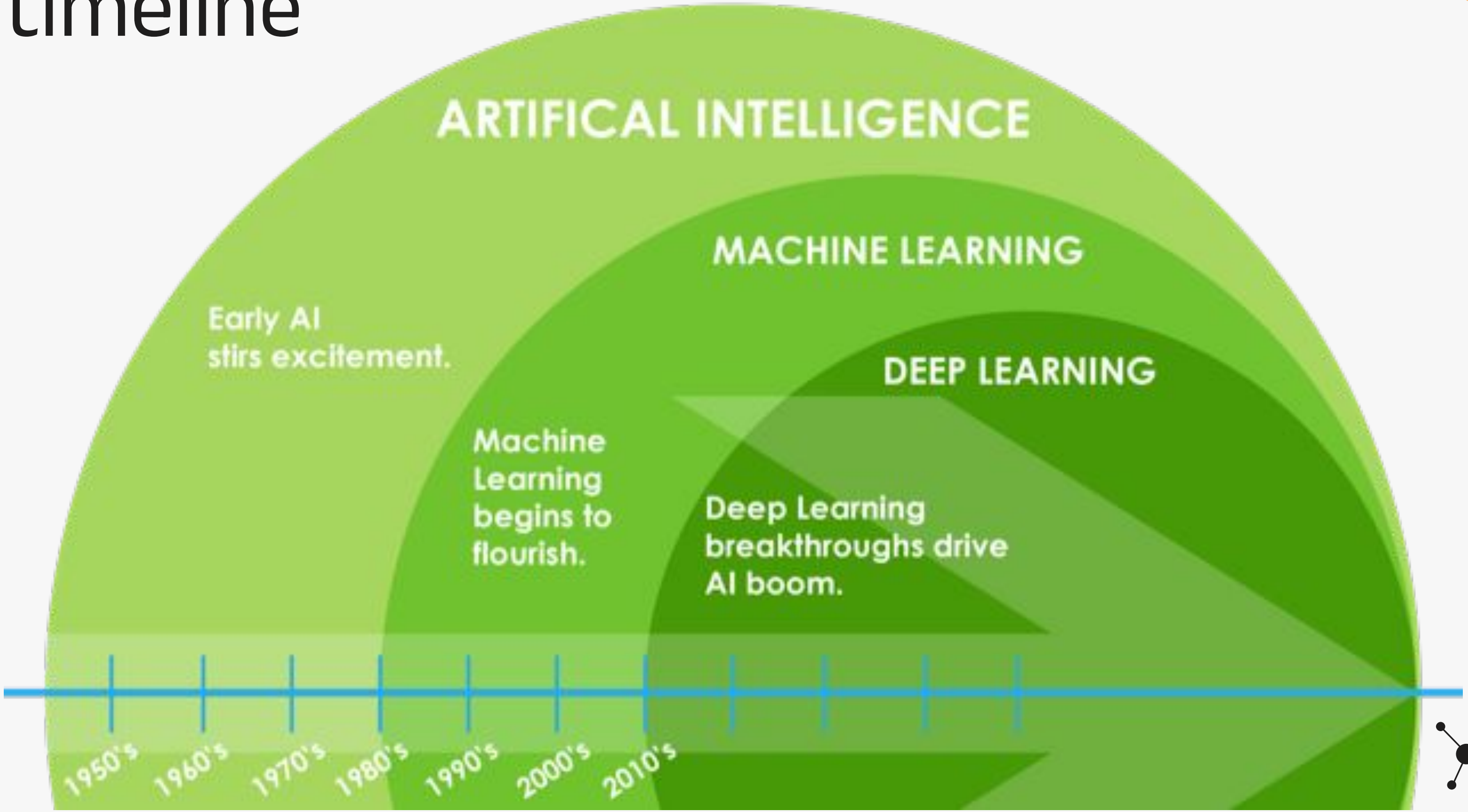
The idea of AI is not new



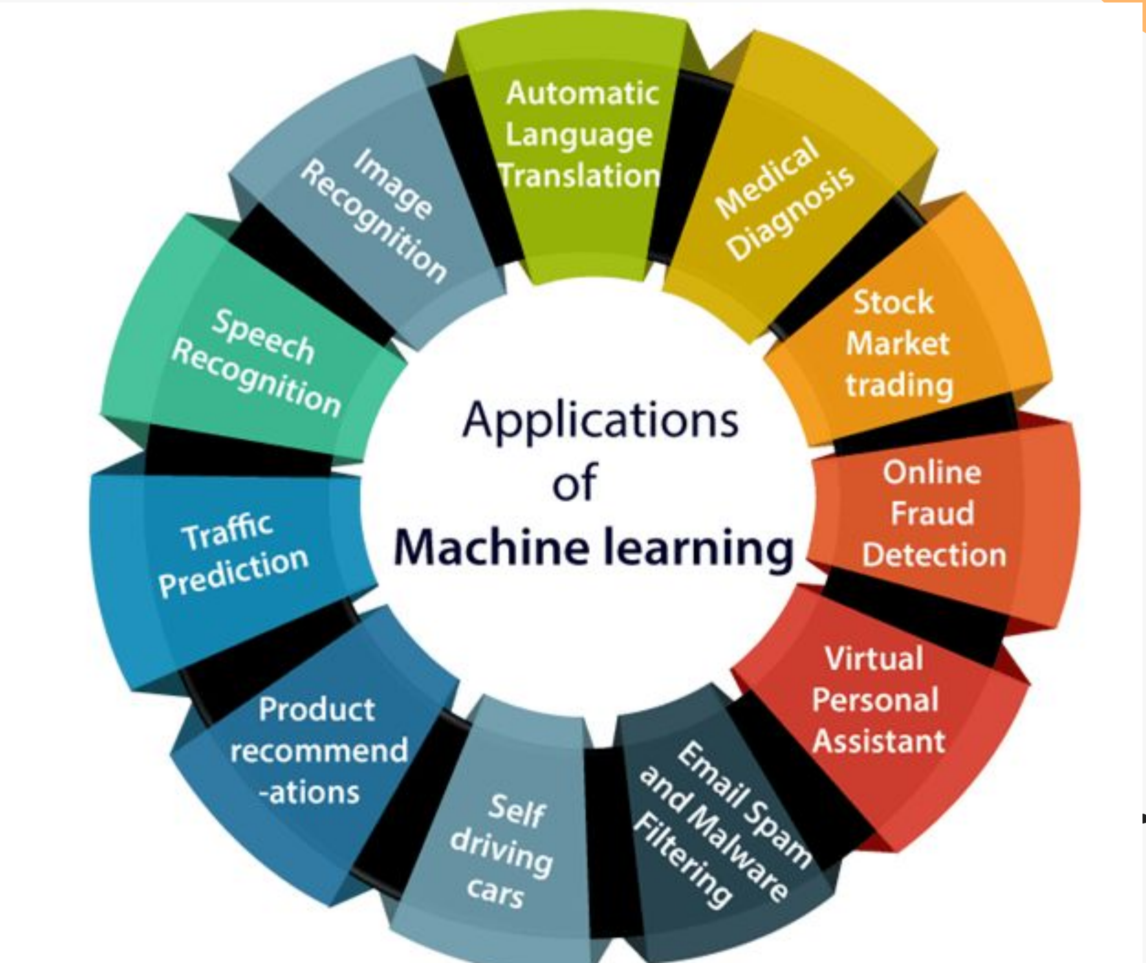
....neither is ML

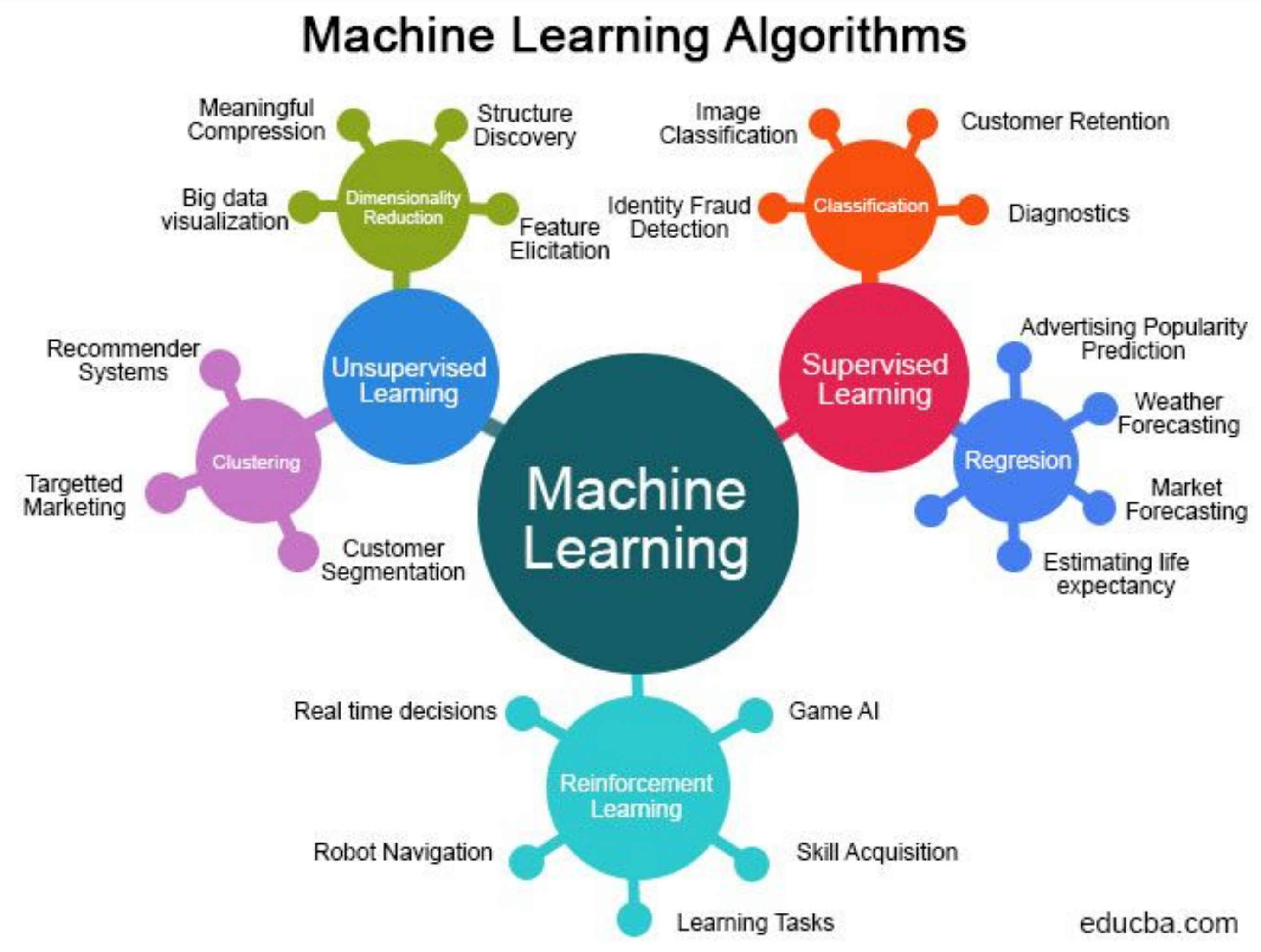
Arthur Samuel demonstrates a checkers playing machine in 1962

AI timeline

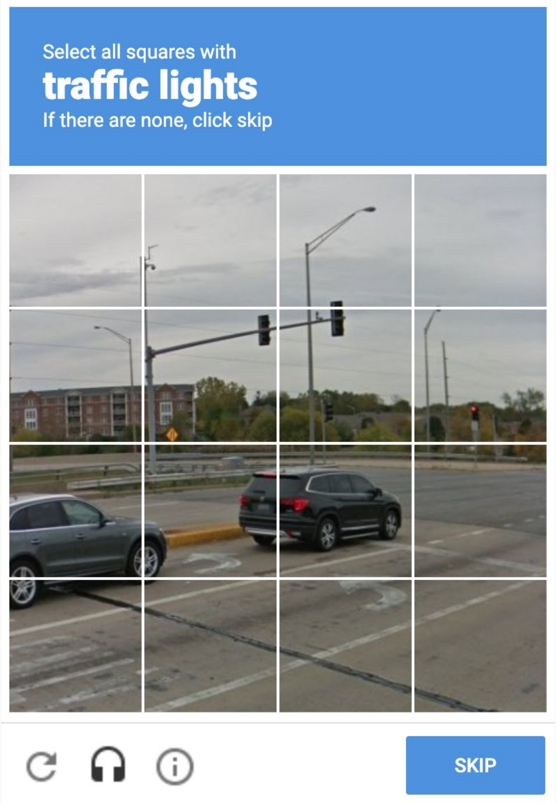

ARTIFICAL INTELLIGENCE
MACHINE LEARNING
DEEP LEARNING
Early AI
stirs excitement.
Machine
Learning
begins to
flourish.
Deep Learning
breakthroughs drive
AI boom.
1950's
1960's
1970's
1980's
1990's
2000's
2010's

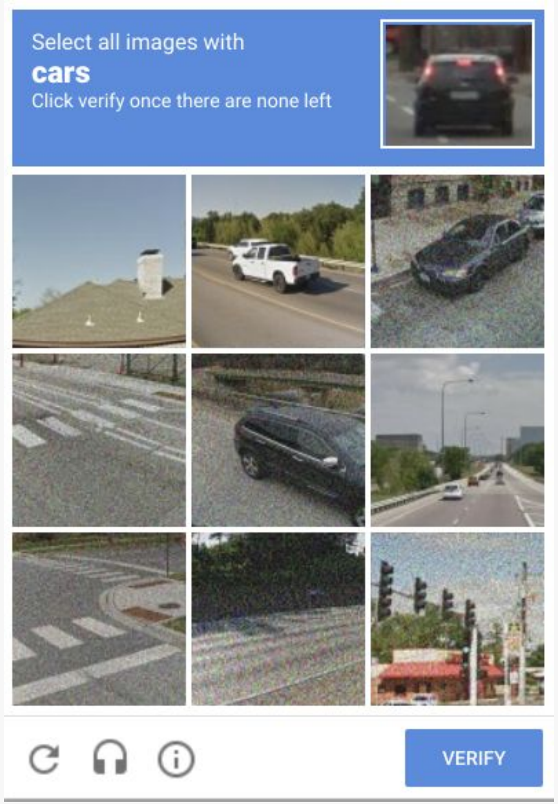
Applied ML

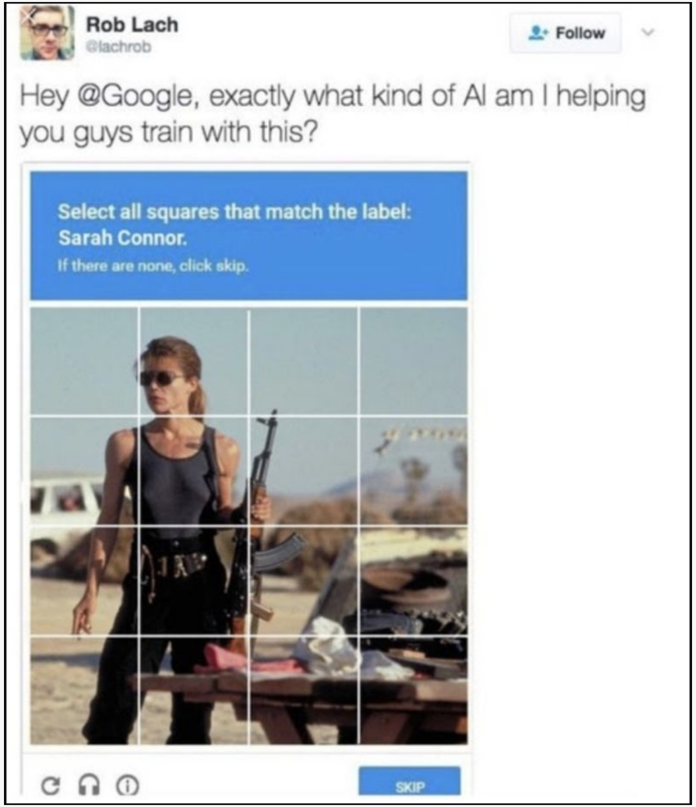

Applications of Machine learning
Automatic Language Translation
Medical Diagnosis
Stock Market trading
Online Fraud Detection
Virtual Personal Assistant
Email Spam and Malware Filtering
Self driving cars
Product recommend -ations
Traffic Prediction
Speech Recognition
Image Recognition


Machine Learning Algorithms
Meaningful Compression
Structure Discovery
Big data visualization
Dimensionality Reduction
Feature Elicitation
Image Classification
Customer Retention
Identity Fraud Detection
Classification
Diagnostics
Unsupervised Learning
Supervised Learning
Advertising Popularity Prediction
Weather Forecasting
Regresion
Market Forecasting
Estimating life expectancy
Recommender Systems
Clustering
Targetted Marketing
Customer Segmentation
Machine Learning
Real time decisions
Game AI
Reinforcement Learning
Robot Navigation
Skill Acquisition
Learning Tasks
educba.com

ML - Supervised Learning

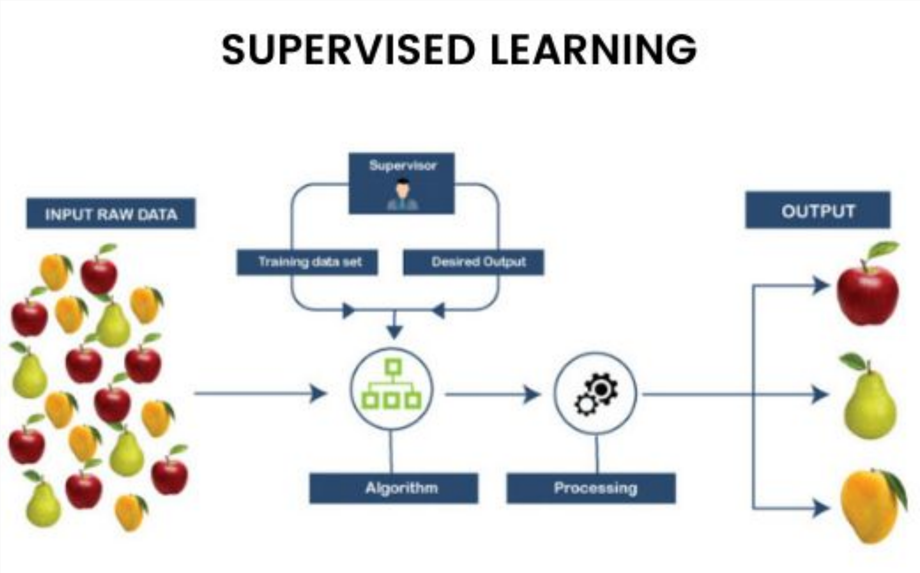


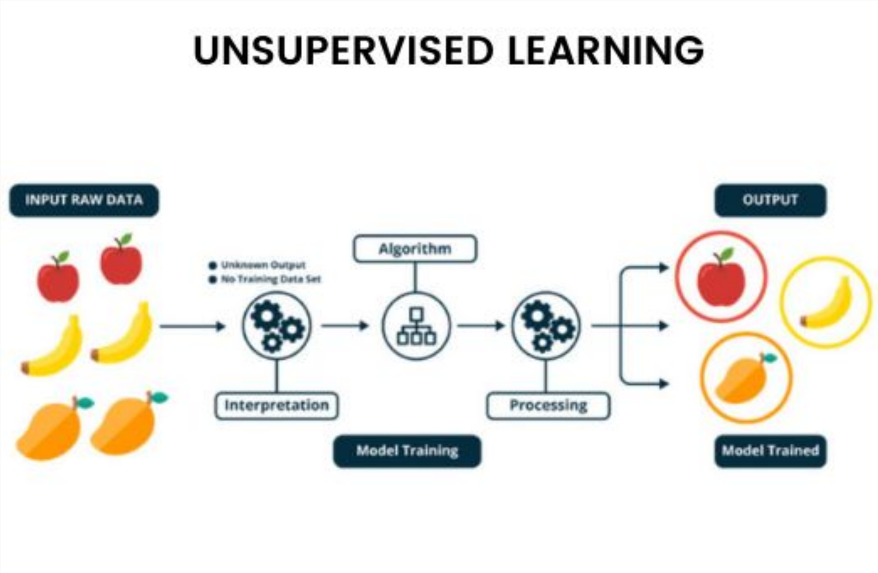




Unsupervised Learning

Introducing the concept of Similarity....!





Self Supervised Learning

Yann LeCun @ylecun

1. Self-Supervised Learning
2. ResNets (not intellectually deep, but useful)
3. Gating -> Attention -> Dynamic connection graphs.
4. Differentiable memory.
5. Permutation-equivariant modules, e.g. multihead self-attention -> Transformers.

> **David Chalmers** @davidchalmers42 · Sep 13, 2022
> what are the most important intellectual breakthroughs (new ideas) in AI in the last ten years?

8:44 PM · Sep 15, 2022

291 Retweets **13** Quotes **1,728** Likes **563** Bookmarks

Meta AI

Research

Self-supervised learning: The dark matter of intelligence

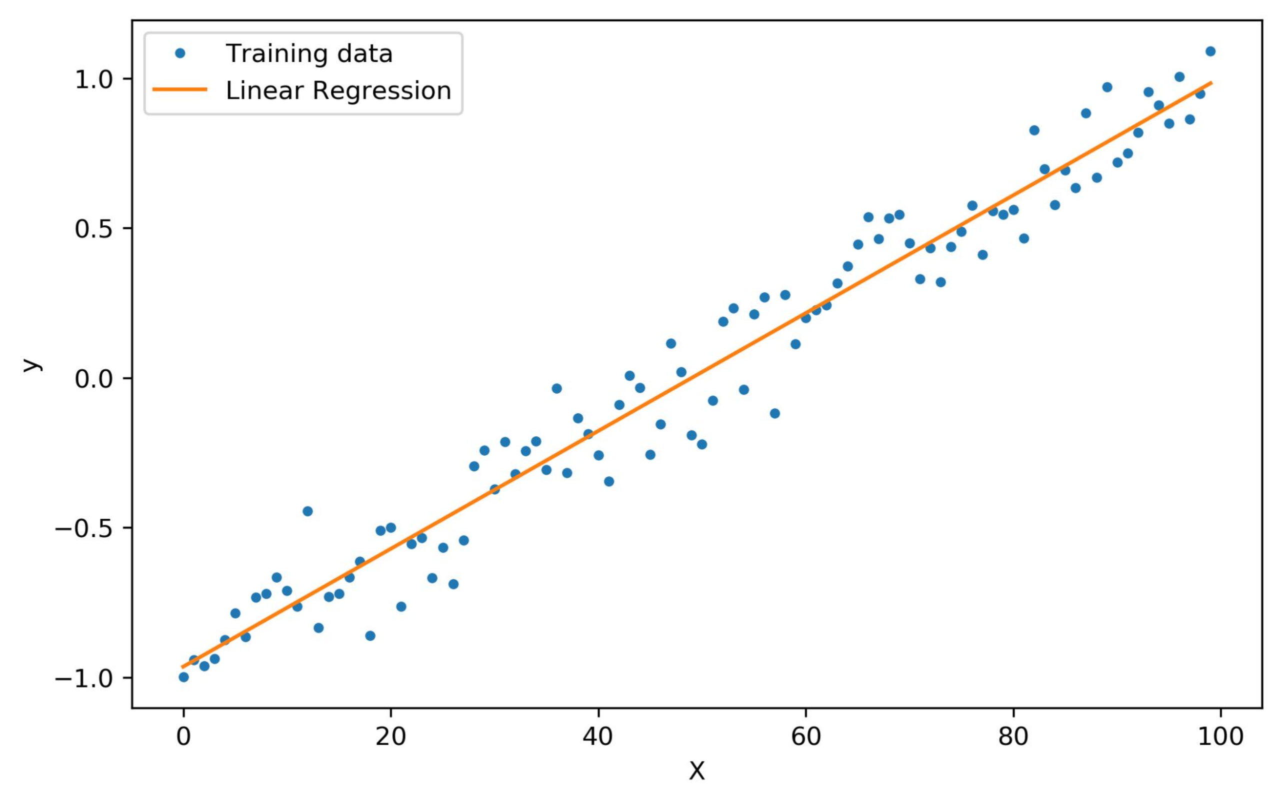
March 4, 2021

In recent years, the AI field has made tremendous progress in developing AI systems that can learn from massive amounts of carefully labeled data. This paradigm of supervised learning has a proven track record for training specialist models that perform extremely well on the task they were trained to do. Unfortunately, there's a limit to how far the field of AI can go with supervised learning alone.

Implicit labels!

What is a model?

What is a ML model?



Parametric Models

Parameters are knobs!
In the linear model we have 2 parameters - slope and intercept
Parametric models can have many parameters - sometimes in the billions! (Like ChatGPT)

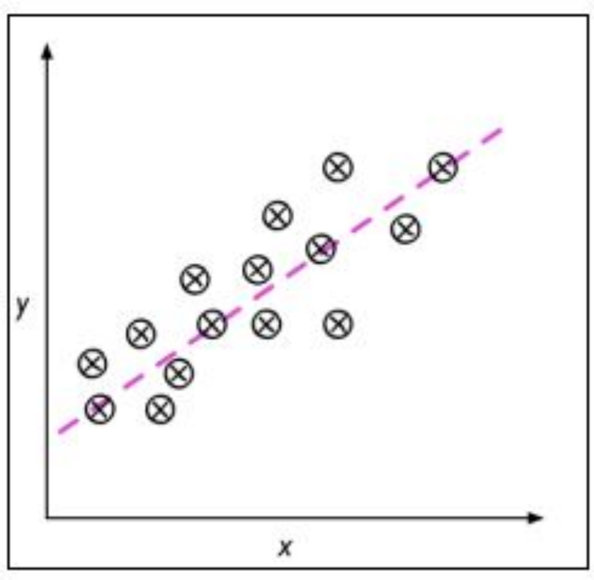
Non-Parametric models like trees will be covered in later slides

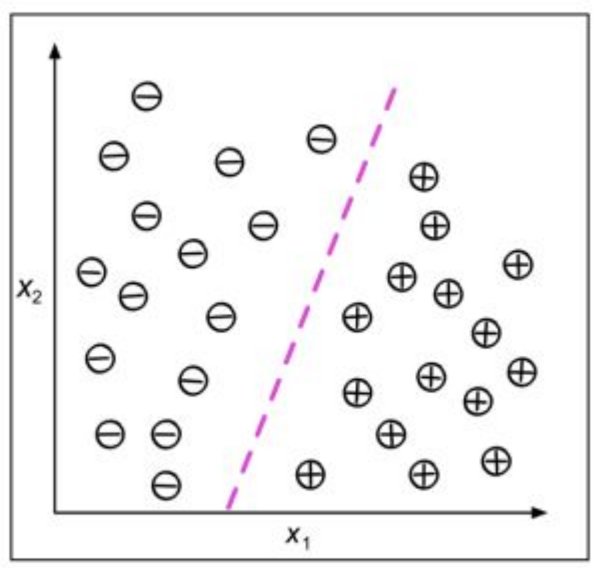
Predicting house prices

Data and Features

	0	1	2	3	4
id	7129300520	6414100192	5631500400	2487200875	1954400510
date	10/13/2014	12/9/2014	2/25/2015	12/9/2014	2/18/2015
price	221900	538000	180000	604000	510000
bedrooms	3	3	2	4	3
bathrooms	1	2.25	1	3	2
sqft_living	1180	2570	770	1960	1680
sqft_lot	5650	7242	10000	5000	8080
floors	1	2	1	1	1
waterfront	0	0	0	0	0
view	0	0	0	0	0
condition	3	3	3	5	3
grade	7	7	6	7	8
sqft_above	1180	2170	770	1050	1680
sqft_basement	0	400	0	910	0
yr_built	1955	1951	1933	1965	1987
yr_renovated	0	1991	0	0	0
zipcode	98178	98125	98028	98136	98074
lat	47.5112	47.721	47.7379	47.5208	47.6168
long	-122.257	-122.319	-122.233	-122.393	-122.045
sqft_living15	1340	1690	2720	1360	1800
sqft_lot15	5650	7639	8062	5000	7503

Regression vs Classification





Setting up the work environment

- Introduction to Jupyter
- Google Colab [https://colab.research.google.com/]

Drawing a line

THE EQUATION FOR THE LINE :

$$y = \underline{mx} + \underline{b}$$

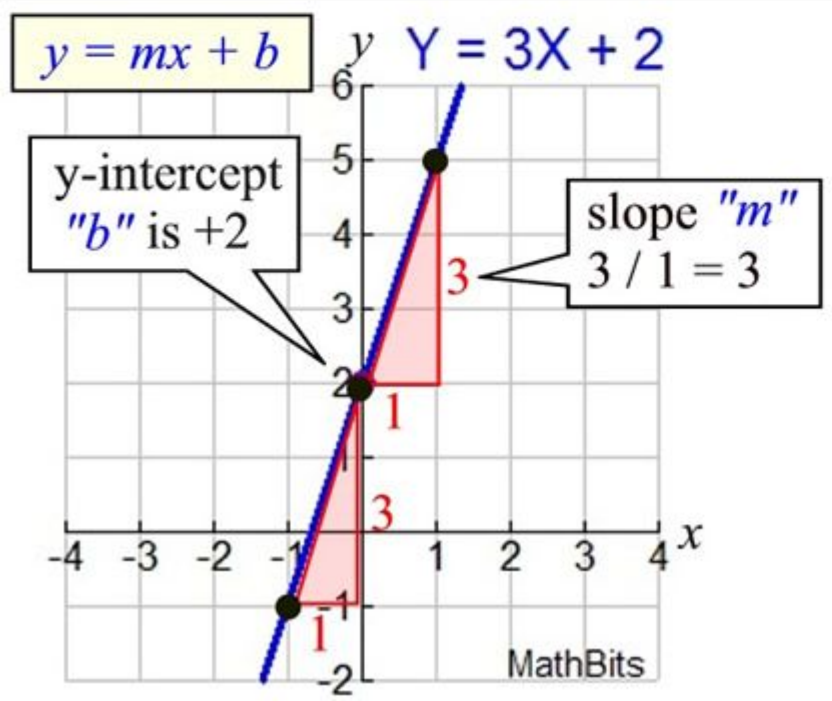
SLOPE

Y-INTERCEPT

$m = 1$
$b = 5$

$\rightarrow$ $y = 1x + 5 = x + 5$

Drawing a line in 2D

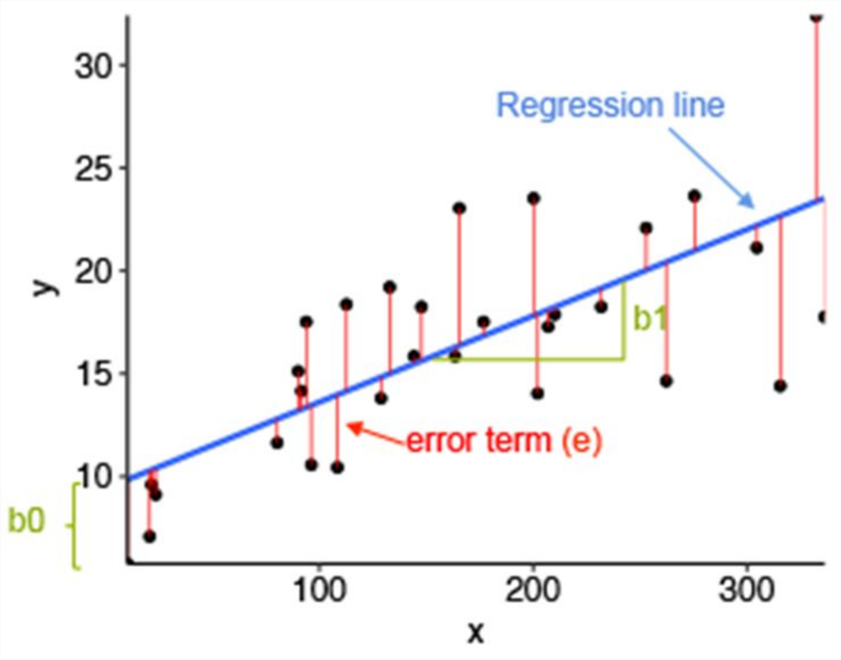


Slope = Rise over run

Let’s look at a live example

— Using desmos

How good is our prediction?

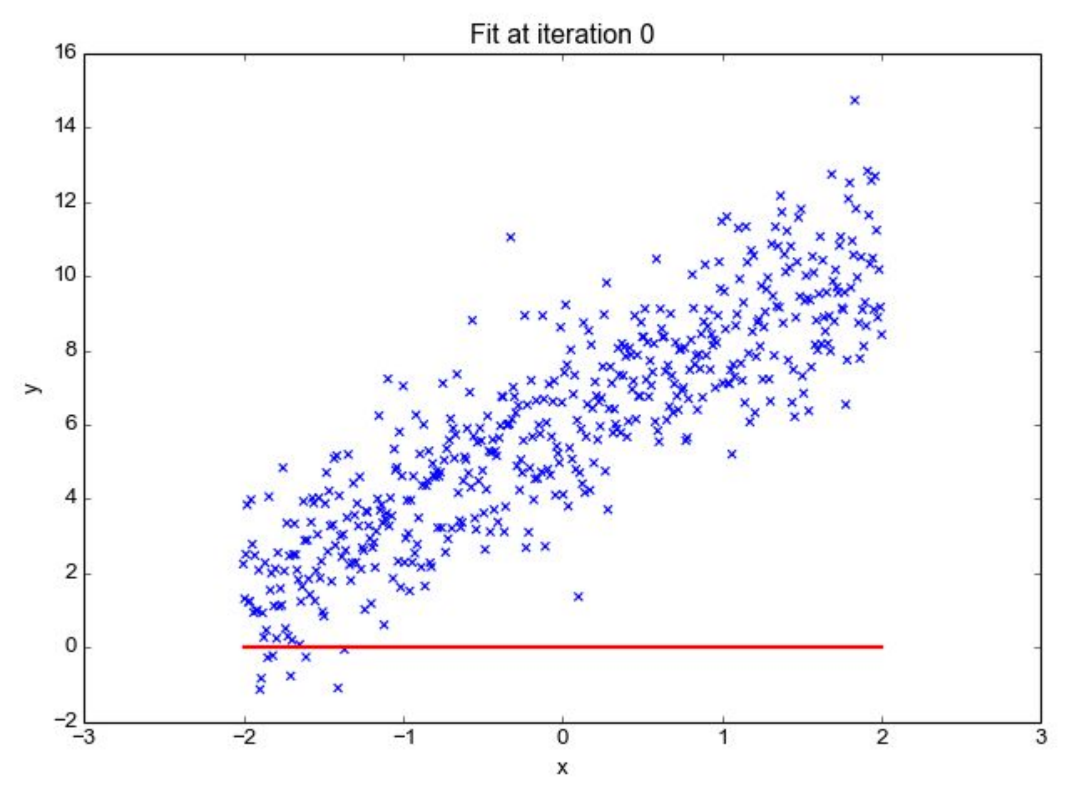


Bad Predictions

$$\mathrm{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2$$

Mean squared error cost function

Training iteratively



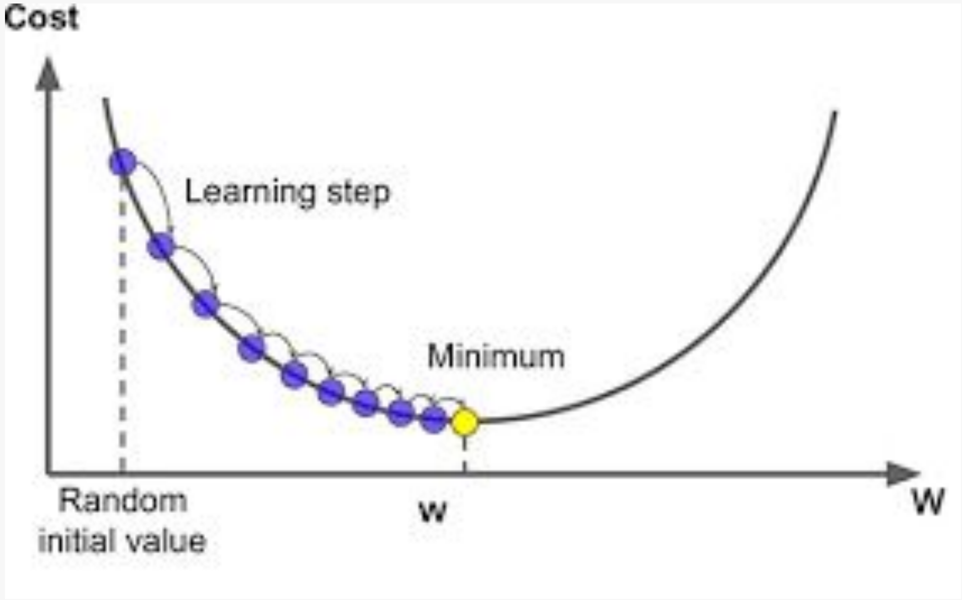
$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2$$

Training a model = Minimizing the loss (or cost)

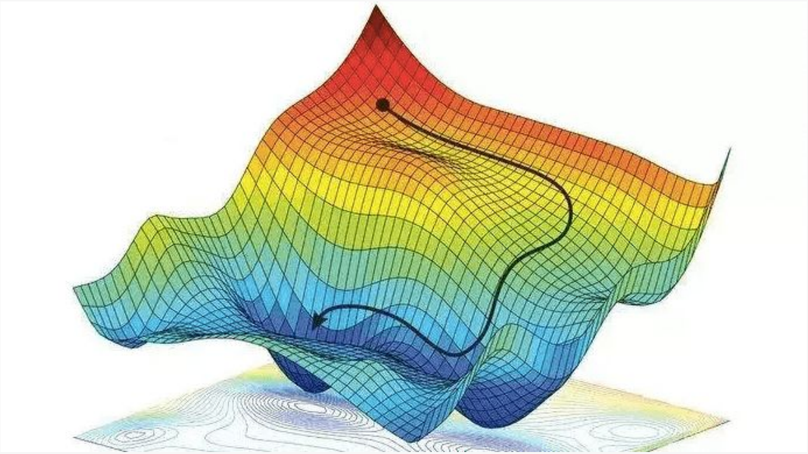
xkcd - ML


THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS

Plotting the loss function



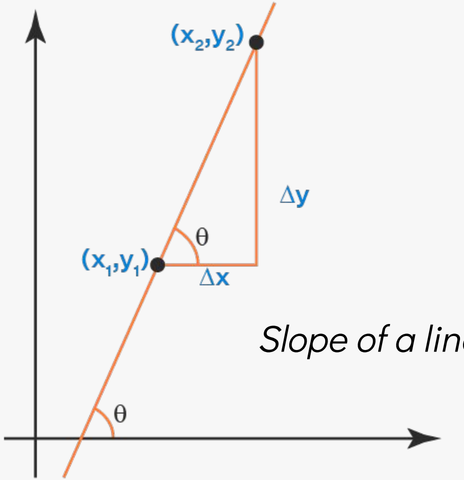
One Parameter MSE loss function



Multi Parameter loss function

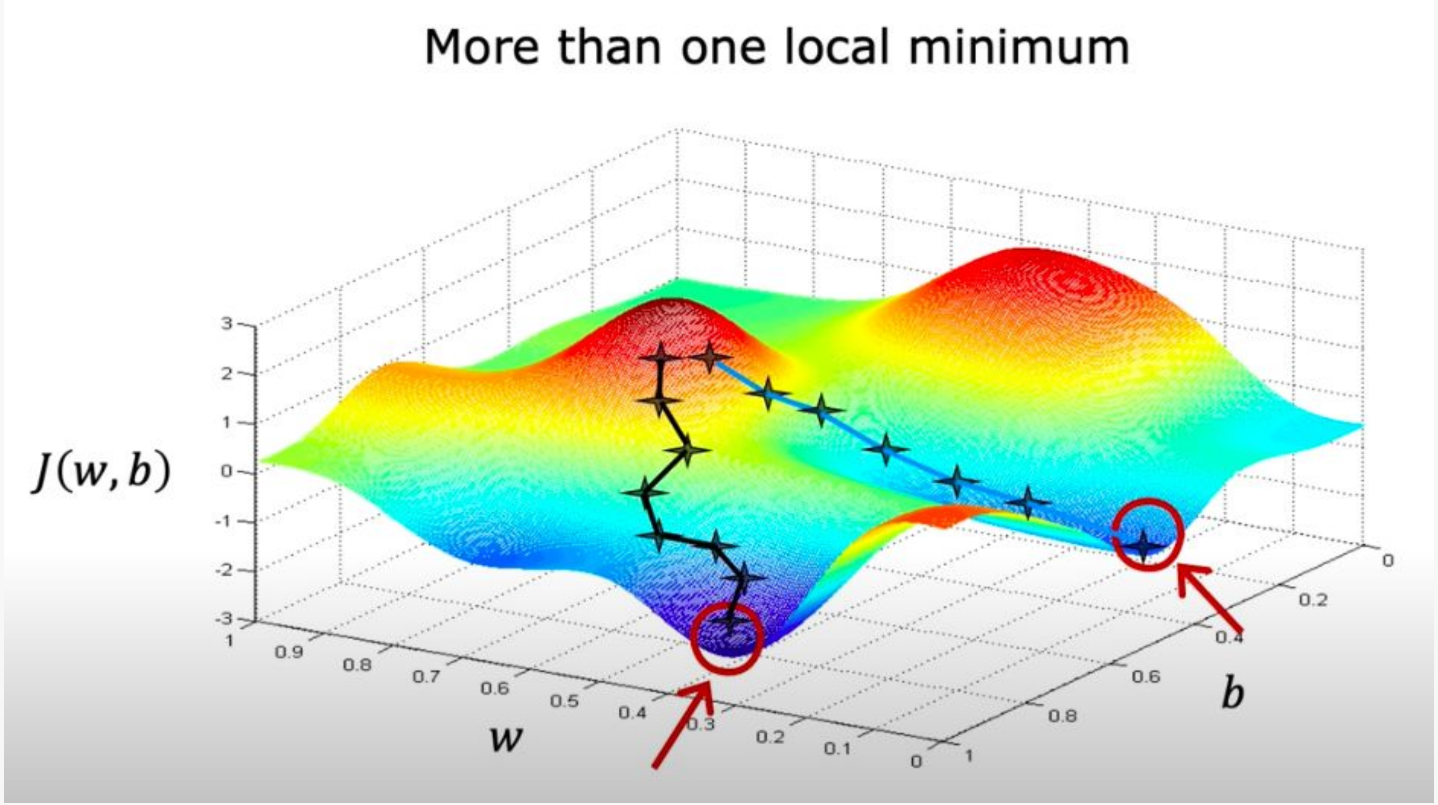
Gradient Descent - the magic behind AI

Using differential calculus to make tiny adjustments to the knobs automatically!

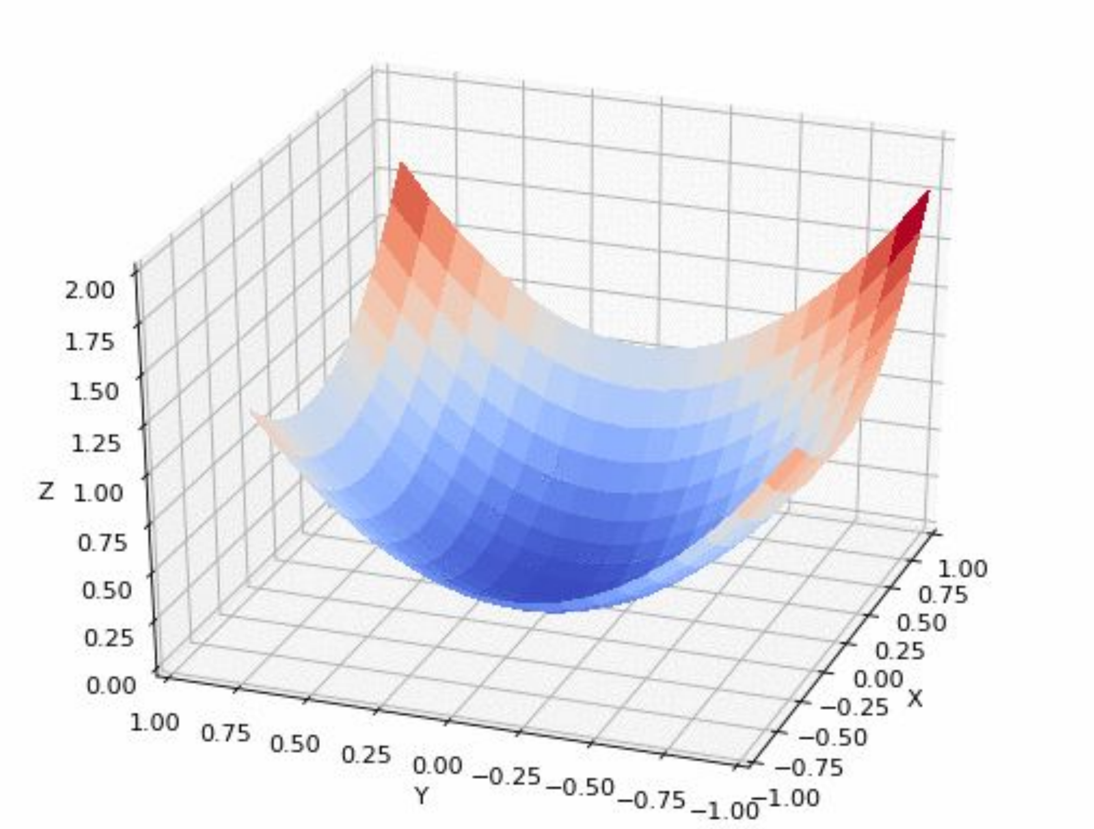


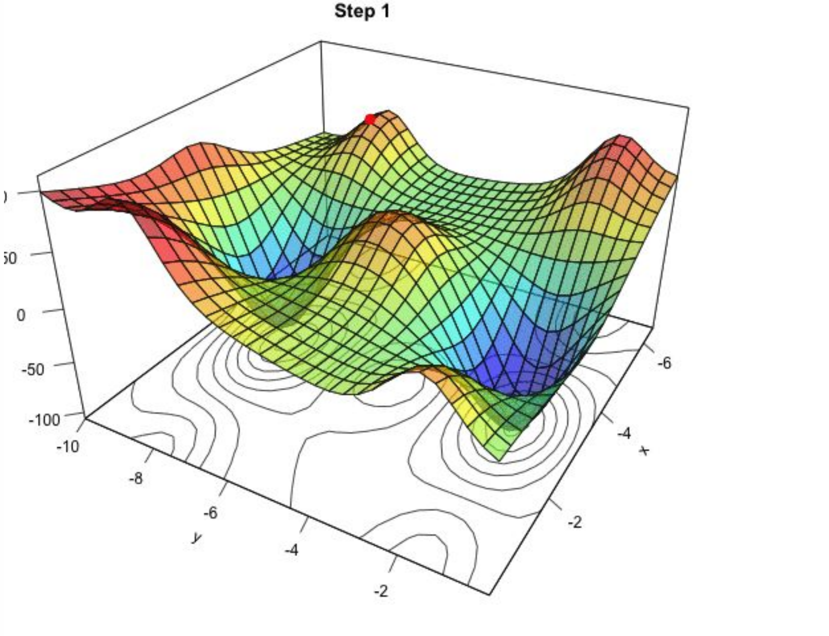
Slope of a line

The loss landscape



Iterative training - step by step adjustments





Gradient Descent - derivatives

Linear regression model

$$f_{w,b}(x) = wx + b$$

Cost function

$$J(w,b) = \frac{1}{2m}\sum_{i=1}^{m}(f_{w,b}(x^{(i)}) - y^{(i)})^2$$

Gradient descent algorithm

repeat until convergence {

$$w = w - \alpha\frac{\partial}{\partial w}J(w,b) \qquad \frac{1}{m}\sum_{i=1}^{m}(f_{w,b}(x^{(i)}) - y^{(i)})x^{(i)}$$

$$b = b - \alpha\frac{\partial}{\partial b}J(w,b) \qquad \frac{1}{m}\sum_{i=1}^{m}(f_{w,b}(x^{(i)}) - y^{(i)})$$

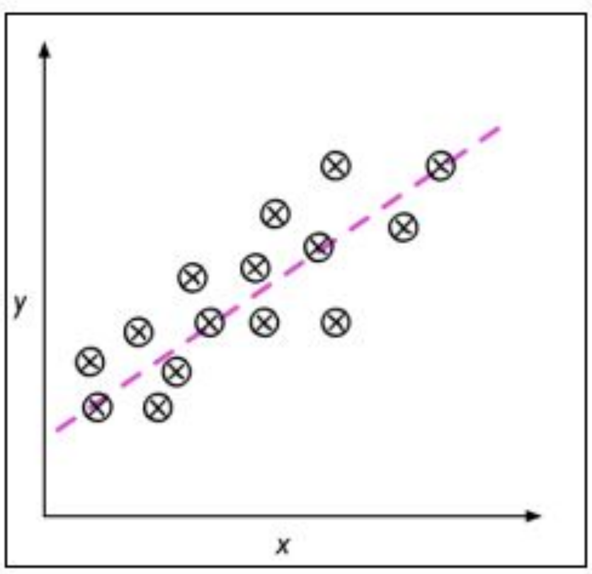
}

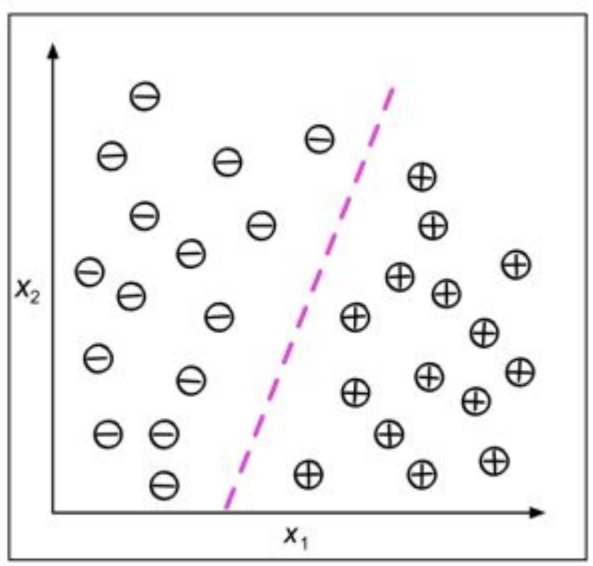
Introduction to Logistic regression

Don't let the name confuse you. For historical reasons it is called 'regression'.

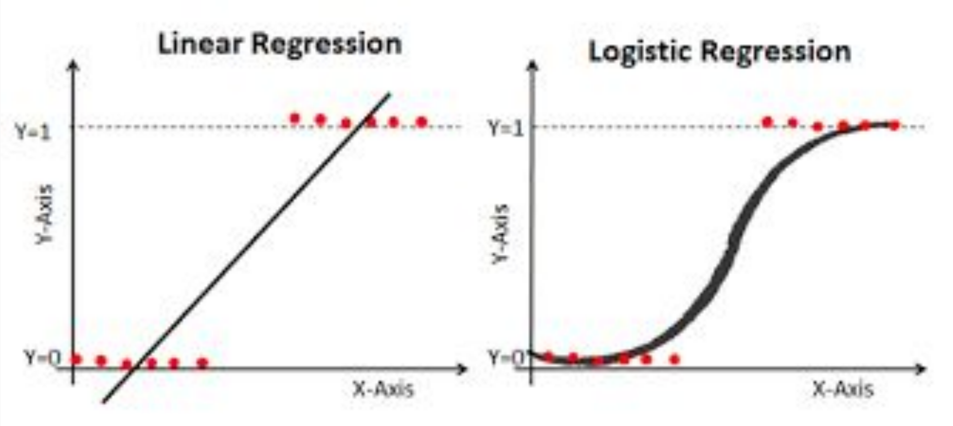
Logistic regression is a *CLASSIFICATION* algorithm!

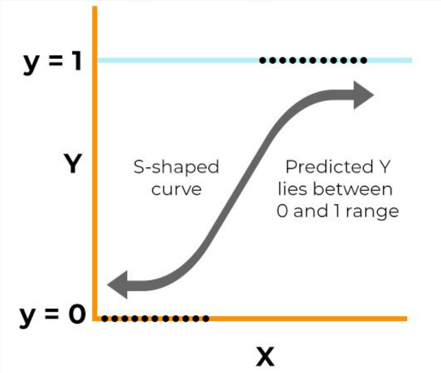
Regression vs Classification



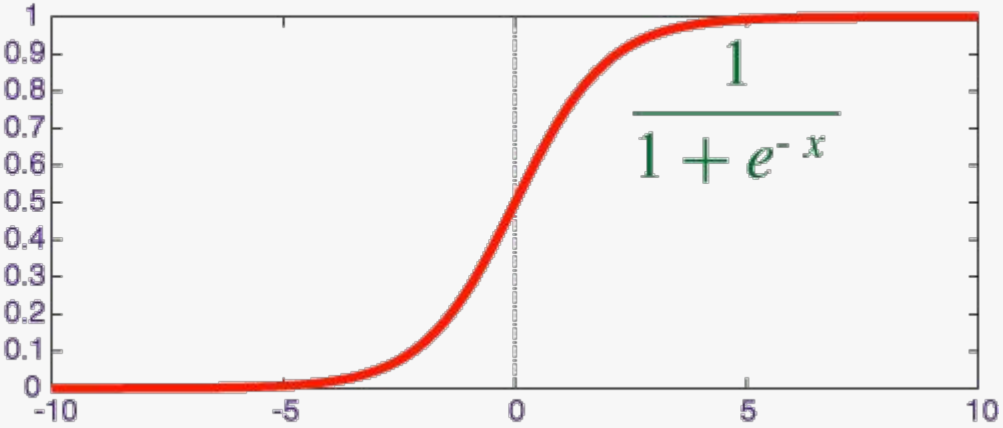


From straight lines to curves

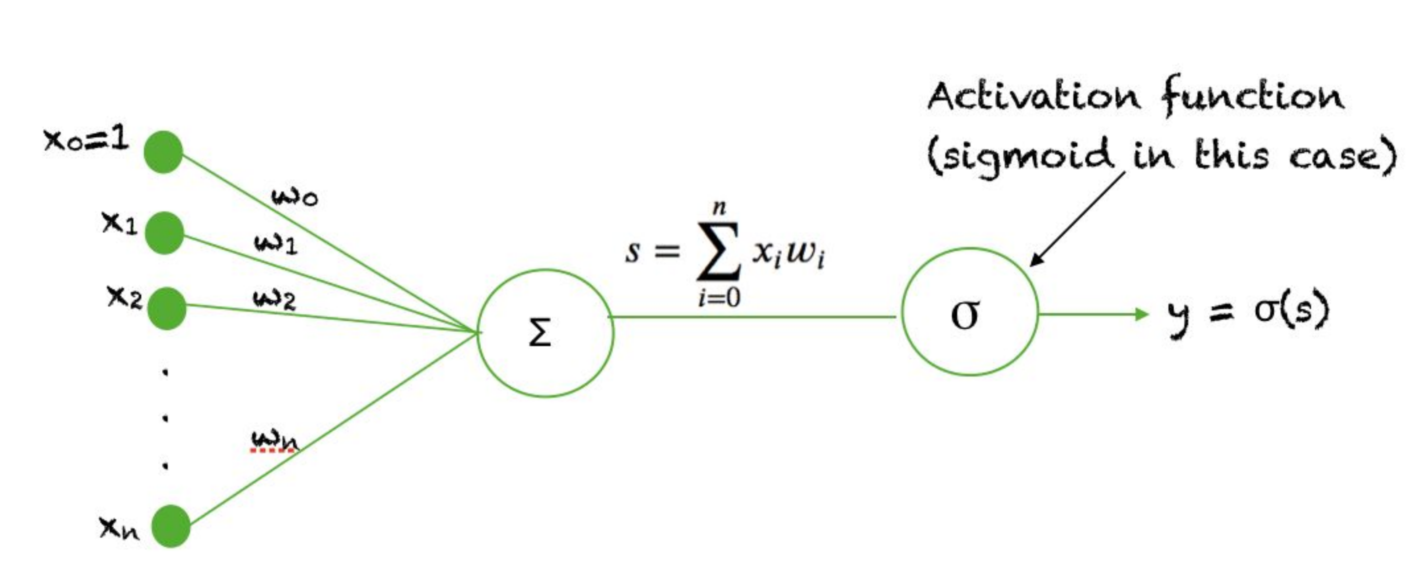




Sigmoid Function



Turning number predictions into class predictions



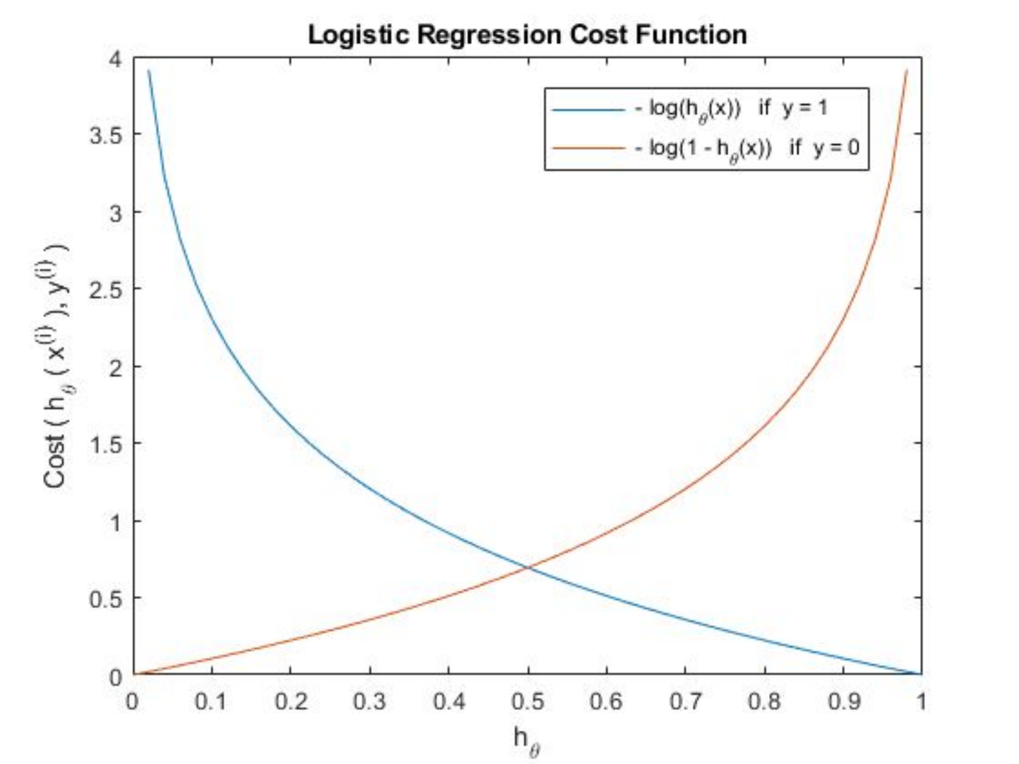
Cost function for logistic regression

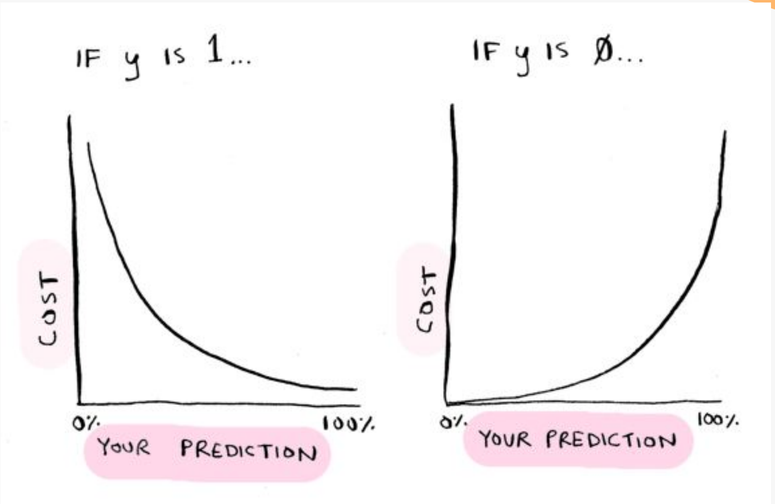
$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} Cost(h_\theta(x^{(i)}), y^{(i)})$$

$$J(\theta) = \frac{1}{m}[\sum_{i=1}^{m} -y^{(i)} log(h_\theta(x^{(i)})) + (1 - y^{(i)}) log(1 - h_\theta(x^{(i)}))]$$

$$m = number\ of\ samples$$

Why logarithm?

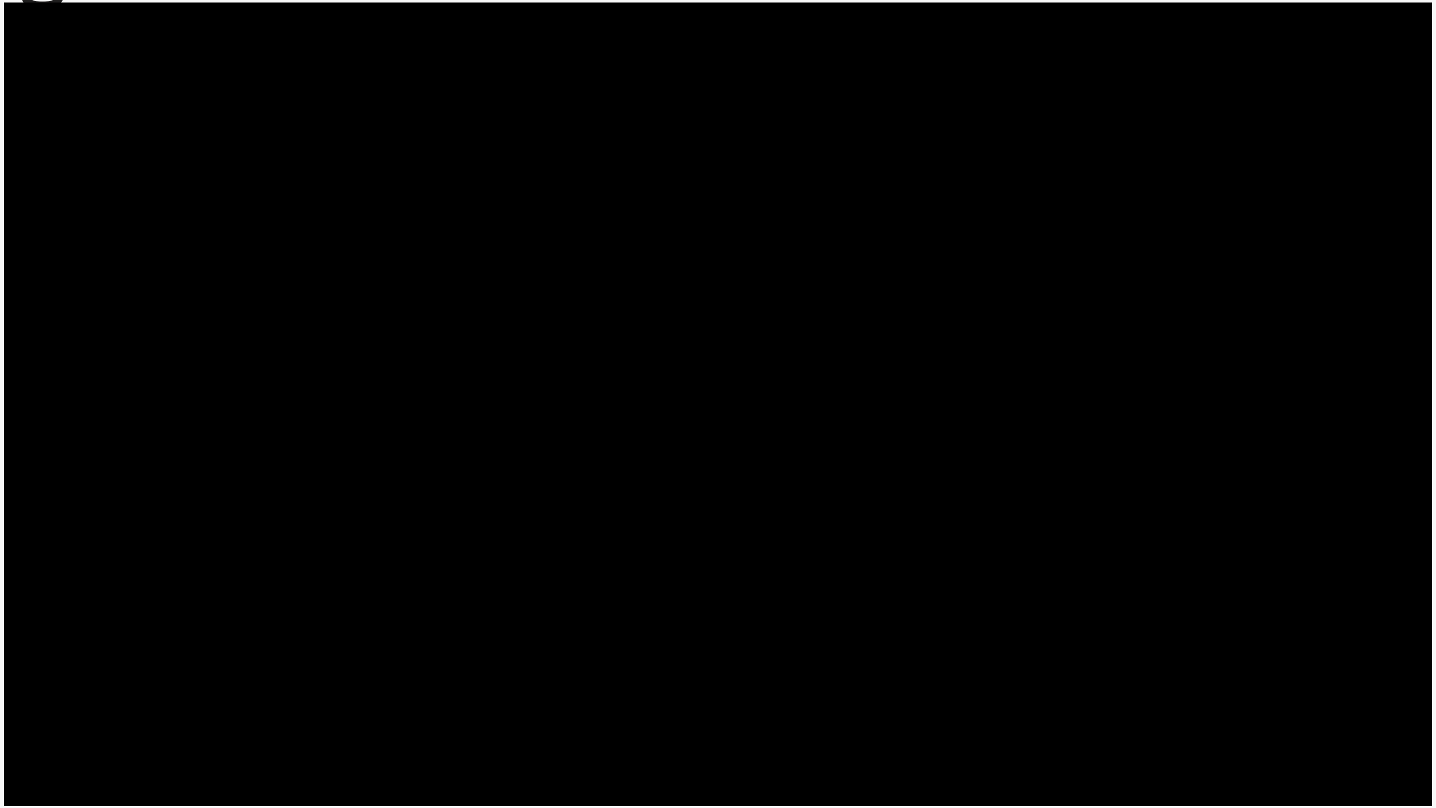




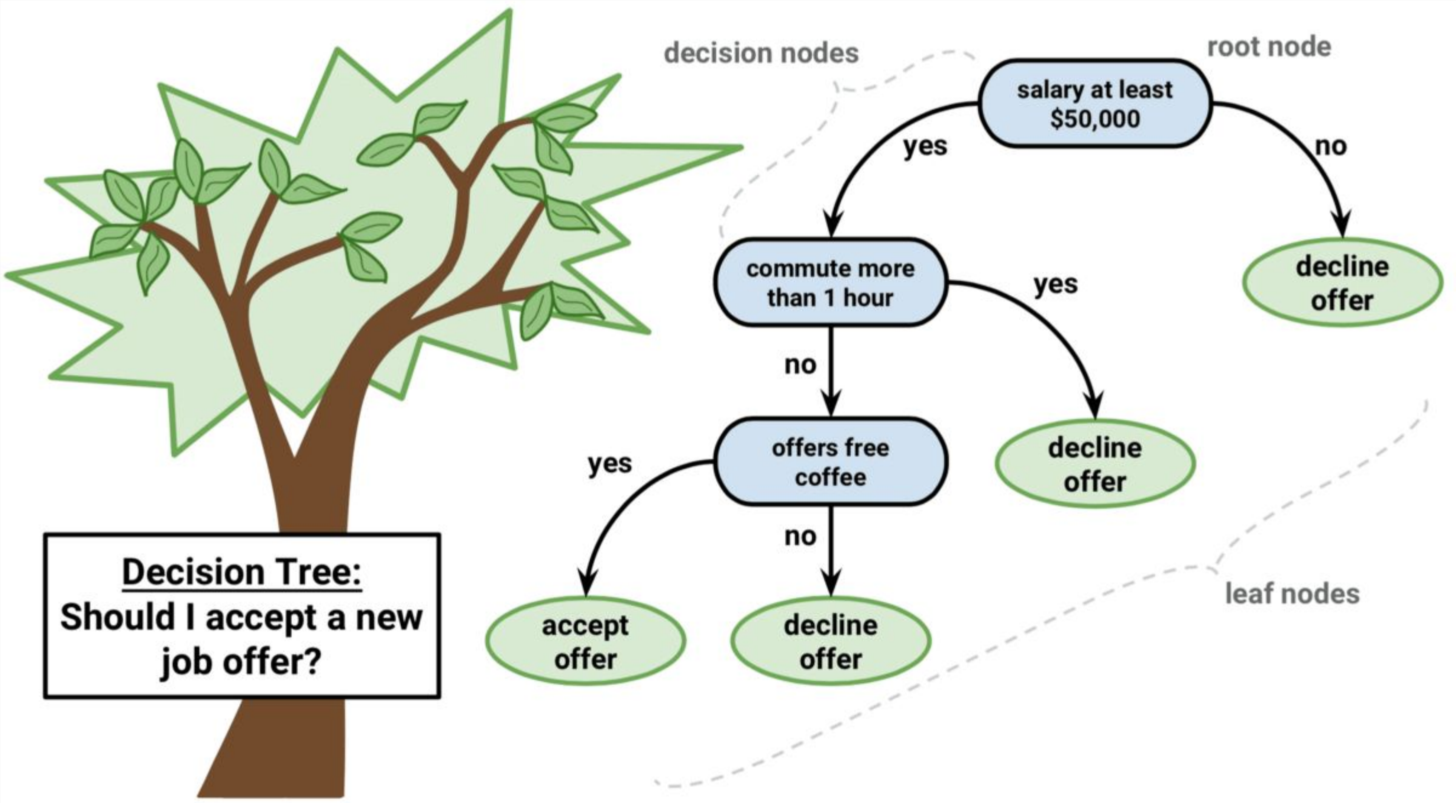
$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} Cost\left(h_\theta(x^{(i)}) - y^{(i)}\right)$$

$$\text{Cost}(h_\theta(x), y) = \begin{cases} -\log\left(h_\theta(x)\right) & \text{if y=1} \\ -\log\left(1 - h_\theta(x)\right) & \text{if y=0} \end{cases}$$

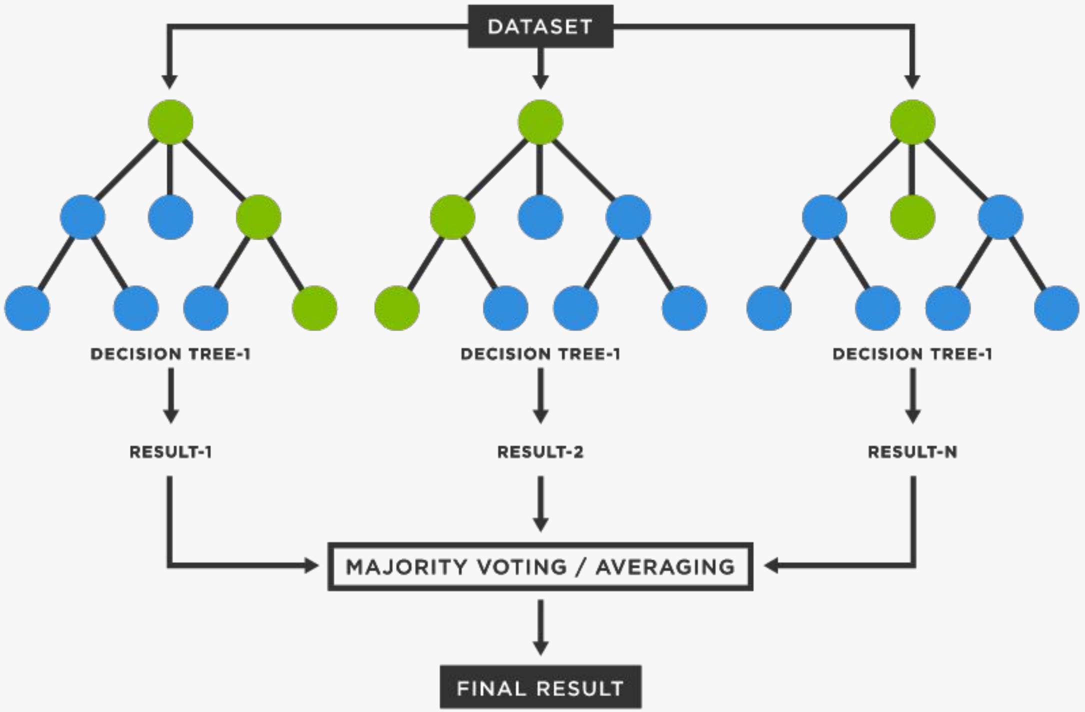
Log loss calculation



Non-Parametric models



Ensemble models



Module 2